DOT/FAA/AM-98/20

Office of Aviation Medicine
Washington, D.C. 20591

An Acoustic Analysis of ATC Communication

O. Veronika Prinzo
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

Philip Lieberman
Emily Pickett
Brown University
Providence, RI 02912

July 1998

Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department
of Transportation
**Federal Aviation
Administration**

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

Technical Report Documentation Page

1. Report No.
DOT/FAA/AM-98/20

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
An Acoustic Analysis of ATC Communication

5. Report Date
July 1998

6. Performing Organization Code

7. Author(s)
Prinzo, O.V.[1], Lieberman, P., and Pickett, E.[2]

8. Performing Organization Report No.

9. Performing Organization Name and Address

[1] FAA Civil Aeromedical Institute
P.O. Box 25082
Oklahoma City, OK 73125

[2] Brown University
Providence, RI 02912

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
95-G-034

12. Sponsoring Agency name and Address
Office of Aviation Medicine
Federal Aviation Administration
800 Independence Ave., S.W.
Washington, D.C. 20591

13. Type of Report and Period Covered

14. Sponsoring Agency Code

15. Supplemental Notes
This work was performed under Task AM-D-96-HRR-513

16. Abstract

This report consists of an acoustic analysis of air traffic control (ATC) communications. Air traffic control specialists (ATCS) from a TRACON facility participated in the simulation study. Each ATCS worked light and heavy traffic density scenarios for 2 feeders and 1 final sector. All communications were audio recorded and transcribed verbatim by a retired ATCS. Workload was determined by the number of aircraft under positive control when the ATCS initiated a transmission. Utterances were selected to achieve maximal workload contrast. For each participant, the 5 lowest workload utterances from the Light version of the scenario (simulating that participant's normal work station) and the 5 highest workload utterances from the Heavy version of the scenario (simulating a work station unfamiliar to the participant) were identified and digitized. For all participants, speaking rate (syllables/second), pause frequency (number of pauses/number of words), and pause duration (duration of pauses/number of words) were generated from the selected utterances using the BLISS speech analysis system (Lieberman and Blumstein, 1988). The results indicate that ATCSs tended to pause more frequently and for greater duration under a light workload condition. The hesitations found in their speech may reflect a shift between a more cognitive "thinking" response mode in light traffic situations where ATCSs know that they have more time to respond and a more automatic mode, which allows them to respond to the increased pace induced by higher traffic loads. In conclusion, it appears that hesitation in speech may be a potential indicator of workload. Despite its highly speaker-dependent nature, hesitation pauses may be a useful indicator of an ATCS's responding in a cognitive, rather than in an automatic mode.

17. Key Words
ATC Workload
Acoustic Analysis
ATC Communications

18. Distribution Statement
Document is available to the public through the
National Technical Information Service
Springfield, Virginia 22161

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 27	22. Price

ACKNOWLEDGMENTS

This research was sponsored by the Federal Aviation Administration's Office of Aviation Safety Analysis (ASY-200), and the Evaluations and Investigations Staff of Air Traffic Services (AAT-20).

TABLE OF CONTENTS

An Acoustic Analysis of ATC Communication

"The true use of speech is not so much to express our wants as to conceal them."

— Oliver Goldsmith (1728-1774)
The Use of Language

1.0 INTRODUCTION

Radio communication is the primary means by which pilots and air traffic control specialists (ATCSs)[1] transmit verbal messages between each other. Controllers learn to speak a particular grammar using a pre-defined cadence during their initial training at the FAA Academy and at their assigned air traffic control facility. In addition to the verbal message transmitted orally, the receiver also receives extralinguistic information conveyed by the speaker. Through additional training and experience, controllers learn to conceal potential emotional content from their speech. While researchers have not yet identified consistently reliable quantifiable factors, several aspects of speech production have been shown to be related to physiological and task-induced stress (e.g., Lieberman et al., 1995; Absil et al., 1995; Benson 1995; Waters et al. 1995; Cummings and Clements, 1990; Frick 1985; Lieberman and Michaels, 1962; Coster 1986; Kagan et al., 1988). The investigation of acoustic correlates of perceived emotional stress and cognitive load is an active field of inquiry in aviation as well (see Prinzo and Britton, 1993 for a review).

The objective of this study was to identify acoustic properties of air traffic control communications associated with changes in workload. Within the context of this study, workload was determined by counting the number of aircraft for which a controller was actively providing radar service when a message was produced. The fewer number of aircraft receiving radar services, the lighter the workload. While ATCSs worked simulated traffic under heavy and light density, their speech was captured onto Digital Audio Tape (DAT) and later analyzed to establish whether previously identified acoustic factors show a reliable relationship to simulated workload.

1.1 Background

Human speech production results from the activity of three functionally distinct systems; (1) the subglottal lungs, (2) the larynx, and (3) the supralaryngeal airway—the supralaryngeal "vocal tract" (SVT). The acoustic consequences of the physiology of these systems have been studied since the early 19th century when Muller (1848) formulated what has come to be known as the "source-filter" theory of speech production. Muller noted that the outward flow of air from the lungs usually provides the power for speech production. If the human auditory system were capable of perceiving acoustic energy at extremely low frequencies, we would "hear" the expiratory airflow. However, the acoustic energy present in the outward flow of air from the lungs is inaudible. The "sources" of acoustic energy for speech are generated by modulating the outward, expiratory flow of air.

Two fundamentally different sources of noise that provide the acoustic energy for the production of human speech are periodic phonation and turbulent noise (Borden and Harris, 1984; Ladefoged, 1962). Periodic phonation is the result of the activity of the larynx. The vocal folds of the larynx, which are extremely complex structures, move inwards and outwards, converting the steady flow of air flowing outwards from the lungs into a series of "puffs" of air. This process repeats itself many times, creating a train of impulses. The number of times the vocal folds open and close per second (i.e., cycles of repetition) directly determines the lowest frequency of the sound that is produced (Sataloff, 1992). Both the basic rate and the detailed airflow through the phonating larynx can be modulated by adjusting the

[1] For ease of reading, the term "controller" will be used synonymously with air traffic control specialist.

tensions of various laryngeal muscles and the alveolar air pressure. The fundamental frequency of phonation (F0) is, by definition, the rate at which the vocal folds open and close. The perceptual response of human listeners of F0 is the perceived pitch of a speaker's voice. Young children, for example, have high F0s during phonation (over 300 Hz); their voices, thus, are "high pitched." The average F0 for men is 125 Hz and over 200 Hz for women. Acoustic energy occurs during phonation at the F0 and at the harmonics of the F0. For example, if F0 is 100 Hz, energy can occur at 200 Hz, 300 Hz, and so on. The amplitude of the harmonics typically decreases as frequency increases for the phonatory patterns typical of human speech. During the course of speech production, speakers constantly modify the fundamental frequency of phonation at linguistic ends (i.e., the speaker's voice will drop at the end of the utterance to signal the listener that the message is complete). Distinctions in dialect, as well as semantic distinctions, can be transmitted by deliberate modifications of the fundamental frequency contour of an utterance. In English, for example, yes-no questions are usually signaled by a rise in F0 at the end of a sentence and stressed words by local peaks in F0 (Lieberman, 1967). For example, "Are you sure?" signals that a question is being asked because of the rise in the F0 when the word "sure" is produced. The response "Yes, I'm sure" has a lowered F0 for the same word occurring at the end of the sentence.

Noise sources tend to have acoustic energy evenly distributed across all frequencies. Noise sources can be generated at constrictions along the airway leading out from the trachea when the airflow becomes turbulent. Noise can be generated at the larynx by forcing air through the partly abducted vocal cords as, for example, at the start of the word "hat." Noise can also be generated by forcing air through constrictions in the SVT. For example, the constriction formed in the mouth when the tongue is raised close to the hard palate in the initial consonant of the word "shoe" generates the noise source of the initial consonant. Momentary bursts of noise excitation typically occur on the release of stop consonants such as [p] when the lips open, at the start of the word "pig." The burst is momentary because the turbulent noise abruptly ceases as the airflow changes from turbulent to laminar flow as the lips open wide.

The time interval between the burst of a stop consonant and the onset of phonation of the following vowel is the voice onset time (VOT). VOT differentiates English "voiced stop" consonants like [b], [d], and [g] from their unvoiced counterparts [p], [t], and [k], respectively. In order to produce a [b], a speaker must initiate phonation soon after opening the lips (within about 20 ms.) to release the pressure in the vocal tract. In contrast, phonation is delayed for 40 ms. or more after lip opening in a [p]. Similar timing distinctions differentiate [d]s from [t]s and [g]s from [k]s. Figure 1 shows the waveforms for a [b] and a [p] produced by the same speaker, where the lip opening (identified by a visible burst) and the onset of phonation (evidenced by periodicity in the waveform) have been marked. The time delay between the marks is the VOT. Normally, speakers of English and many other languages maintain the distinctions between voiced and unvoiced stop consonants by keeping the VOT regions of the two separated by at least 20 ms.

1.2 Measures of Interest

Four primary measures of interest were selected as dependent variables: (1) speaking rate, (2) hesitation, (3) fundamental frequency (F0), and (4) voice onset time (VOT). Speaking rate (syllables/second) might covary with workload in either of two directions. It is possible that, because an increase in work load requires an increase in the number of communications in a fixed amount of time, speaking rate would be increased to "squeeze in" more information in a given time period. Conversely, it might be the case that, as workload increases, speaking rate decreases. It has been shown that verbal "hesitation," which is typified by brief silence, increases with task difficulty and with the quality of a cognitive solution to a given task (Eisler 1968). It has further been shown that there is an inverse relation between the amount of hesitation and speaking rate (Eisler 1968). Assuming increasing workload is equivalent to increasing task difficulty, speaking rate may decrease as workload increases.

In light of Eisler's findings, hesitation was determined as a measure of potential interest. Eisler established that in general, 40%-50% of speech is actually silence; that is, speech is not the continuous flow of sound indicated by our perception. Three types of silence can be found in connected speech: 1) the

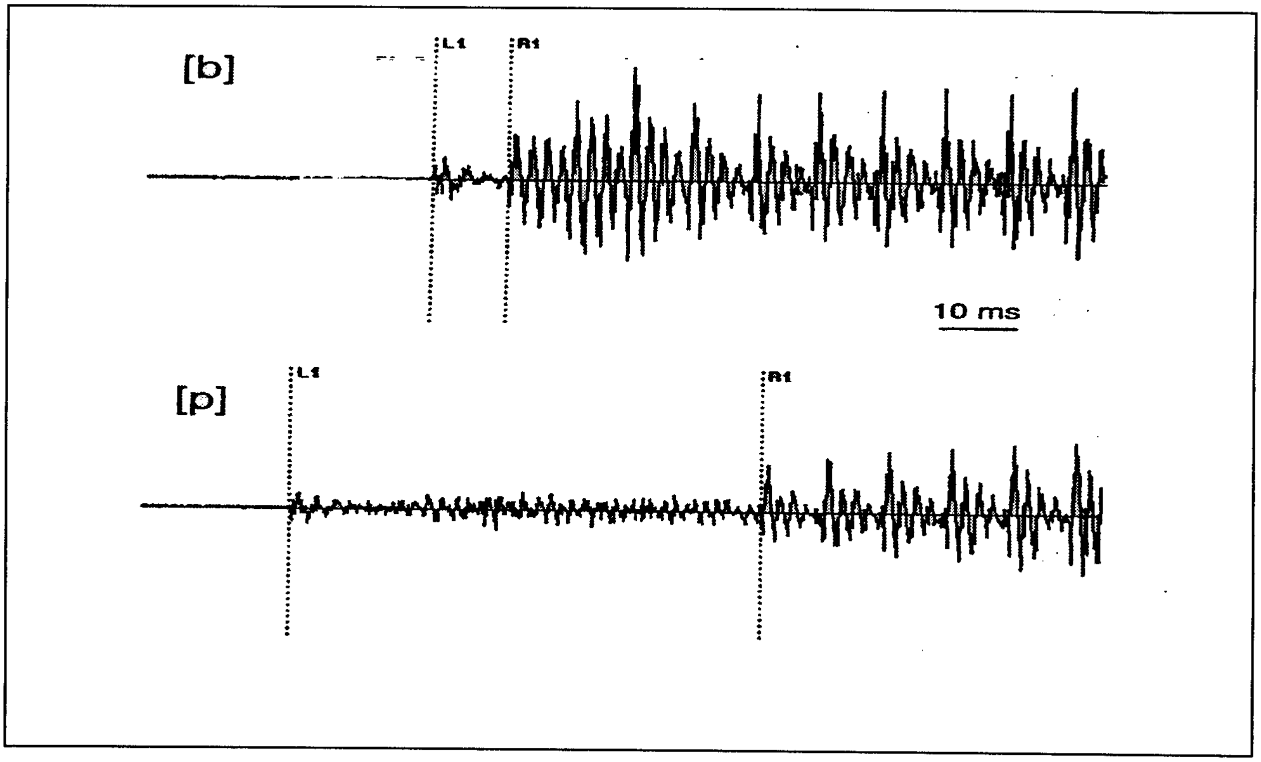


Figure 1. Speech waveform segments corresponding to a [b] and a [p] spoken by the same speaker under identical conditions. Cursors have been placed at the onset of the burst that was caused by opening the lips (L1) and at the onset of periodicity that indicates vocal fold vibration (R1). The marked interval, Voice Onset Timing (VOT) is used by speakers and listeners to differentiate the two types of consonants in word-initial positions.

discontinuity of phonation that occurs in articulatory shifts, e.g., when two stop consonants follow each other; 2) discontinuity of phonation attributable to hesitation; and 3) the gap in speech required for inhalation. The second type of silence, hesitation pauses, is associated with complexity of general planning, task difficulty and the quality of a cognitive solution. Eisler and her colleagues showed that a person making reasoned responses to a question had longer type 2 pauses than a person responding in an automatic, non-engaged manner. Pause duration thus can reflect "thinking time." Therefore, increased hesitation may be due to an increase in pause frequency, pause duration, or both. Accordingly, we used 2 measures to examine hesitation; number of pauses per word and average duration of pauses per word. Again, if an increase in workload results in an increase in cognitive load or thinking time, hesitation, as reflected in either measure, may increase.

Although it is generally accepted that fundamental frequency (F0) is affected by physiological and psychological stress, there is conflicting evidence as to which specific properties of F0 are involved (e.g., Lieberman et al., 1995; Absil et al., 1995; Benson 1995; Waters et al., 1995; Cummings and Clements, 1990; Frick 1985; Lieberman 1963). We selected a global and a local measure: the overall pitch contour of an utterance, and the pitch period[2] of the highest amplitude portion of the second vowel in the word "approach" when it occurred in the final segment of an utterance, that is the utterance-finally.[3]

VOT, measured as the time interval between the burst stop consonant and the onset of phonation, is an objective acoustic measure of speech production.

[2] The pitch period is also commonly referred to as fundamental frequency determination.

[3] Often, the judgment of pitch for aperiodic sounds is influenced by the frequency measured by hertz, the number of cycles per second (Hz), at which the amplitude is highest.

It reflects a participant's ability to precisely sequence the maneuvers of the tongue, lips, velum and larynx that are necessary to produce human speech. Studies of Broca's aphasia (Blumstein et al., 1980; Baum et al., 1990), Parkinson's Disease (Lieberman et al., 1992) and mountain climbers breathing low oxygen-content air in the course of an ascent of Mount Everest (Lieberman et al., 1995) show that control of VOT deteriorates. In these cases, abnormal VOT production is correlated with decrements in reasoning and sentence comprehension. As such, it has been suggested that VOT production may be used as an index of cognitive functioning. Accordingly, we measured the VOT of the word-initial voiceless velar stop [k] and the word-medial voiceless alveolar stop [t] from the word 'contact' when it occurred utterance-finally in the phrase 'contact approach' (26% of utterances). This data set was chosen to minimize variation attributed to context, both lexical and phrasal.

2.0 METHODS

2.1 Participants

Twelve full performance level (FPL) air traffic control specialists from a level 5 Terminal Radar Approach Control[4] (TRACON) facility completed this study. There were 5 East specialty and 7 West specialty air traffic controller specialists (9 male and 3 female) who, collectively, had 13.17 mean years of terminal experience (SD = 3.49) with 9.88 mean years (SD = 3.19) at the full performance level. The East specialist only works sector positions that provide radar services to aircraft arriving from the east and the west specialist only works sector positions that provide air traffic services to aircraft arriving from the west.

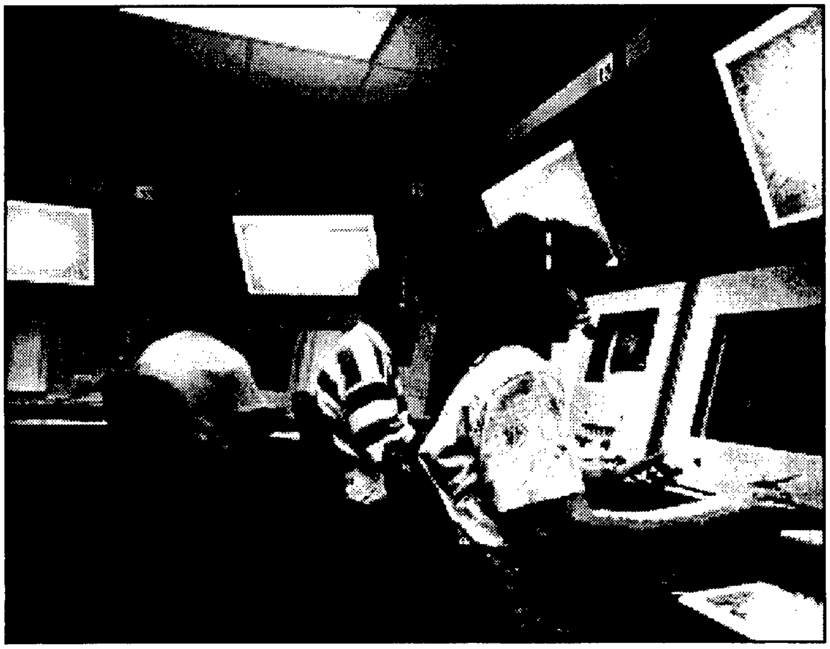


Figure 2. TRACONpro simulator.

2.2 Equipment

2.2.1 TRACON and Ghost Pilot Workstations. Wesson International's TRACONpro© software was installed on two 486/66 MHz DX2 personal computers. Each workstation displayed radar traffic on a 21" multi-scanning capable monitor with high-resolution video adapters (1280 x1024x256). As shown in Figure 2, the TRACON workstation included an amber 14" monitor for displaying automatic terminal information service[5] (ATIS), a track ball, and automated radar terminal system[6] (ARTS IIIA) simulated keyboard, standard 101-style keyboard, Verbex 6000 Voice Systems continuous voice recognition "slave" computer board, push-to-talk headset, and Soundblaster 16-bit digitized pilot response sound board. The ghost pilot workstation included a

[4] A terminal radar approach control (TRACON) facility is associated with an air traffic control tower that uses radar to provide approach control services to aircraft.

[5] Automatic Terminal Information Service provides pilots with continuous broadcast of recorded nonradar information in selected terminal areas. Information includes time, weather, runway, and other essential but routine information. This information is displayed on a secondary monitor next to the radar display.

[6] The Radar Tracking and Beacon Tracking Level of the modular, programmable automated radar terminal system. ARTSIIIA detects, tracks, and predicts primary as well as secondary radar-derived aircraft targets. This more sophisticated computer-driven system upgrades the existing ARTS III system by providing improved tracking, continuous data recording, and fail-safe capabilities.

standard 101-style keyboard and computer mouse. The TRACON workstation was housed in a room separate from the ghost pilot workstation. The workstations communicated to each other though a LANtastic network operating system.

2.2.2 Video Recording Equipment. A Sony Handycam CCD-TR81 video Hi8 camcorder, mounted on a Bogen 3165 Tripod, was positioned approximately 4 meters to the left and 6 meters in front the controller's workstation. Only the radar display, back of the controller, and hand movements were recorded. The audio/video output of the Sony Handycam went to a 3-set Audio/Video Distribution Amplifier (15-1103), displayed on a Sony Color Video Monitor PVM2530 equipped with 2 Sony SS-X6A speakers, and recorded by a Sony Video Cassette Recorder SVO-1610 on standard VHS T120 Cassettes.

2.2.3 Audio recording equipment. A Sony Electret Condenser Microphone (ECM-77B) was attached to a Shurlite headset and positioned approximately 1.5 cm from the controller's lips. The output signals of the microphone were amplified by a Panasonic Audio Mixer WR-450 and then sent to a Sony Digital Audio Recorder PCM-2700, where they were time stamped and stored on 120-minute BASF DAT Cassettes.

2.3 Technical Support Staff

A certified human ghost pilot from the FAA Academy was trained on the 6 scenarios and served as the ghost pilot in this study. A recently retired FPL controller served as the subject matter expert. He constructed the scenarios, trained the ghost pilot, developed briefing materials, and provided the ghost pilot with on-line instructions while the controller worked the scenario. Several staff members from the TRACON facility provided expert information and guidance in the development of the airspace, procedures, and traffic. Also, several controllers worked the scenarios at the workstation prior to the start of the experiment, reviewed each scenario, and provided guidance to ensure fidelity and realism.

2.4 Materials

2.4.1 Scenario Construction. The number of aircraft requiring radar service was experimentally manipulated to simulate high and low workload scenarios. For example, light traffic density involved approximately 1 aircraft communicating with the ATCS per minute and heavy traffic involved 2 aircraft communicating with the ATCS per minute. Light traffic scenarios were developed from heavy traffic scenarios by simply removing 50% of the aircraft from the scenario. The Feeder[7] East, Feeder West, and Arrival[8] positions were simulated. For example, the East specialist will never work on the West side. Traffic density was crossed with simulated positions to produce 6 scenarios.

2.4.2 Ghost Pilot Communication Scripts. Based on analyses performed by Prinzo (1996) on ATC/pilot voice communications acquired from the participating TRACON facilities, normal and problematic pilot communication scripts were constructed and fully counter-balanced for use in each scenario. The scripts were used by the ghost pilot, who initiated calls to ATC at pre-determined times and responded to messages generated by the controller.

2.4.3 Computer-Generated Pilot Responses. Each Non-Target aircraft response was generated by the TRACONpro software. Aircraft call signs, ICAO alphabet, and phrases used in operational communications were recorded, edited, and stored as .WAV files. The intelligibility and realism of the computer-generated responses was evaluated by the FBI speech-processing laboratory at Quantico, VA. A computer-generated response was selected at random and compared with the live recording of that message by the ghost pilot. A visual inspection of the spectrograms revealed that the visual characteristics of the sound waves were the same and produced by the same person.

2.5 Procedure

Upon arrival to the TRACON simulation laboratory on Day 1, the controller was briefed on the purpose of the study, instructed on Verbex voice

[7] A Feeder sector is a transition area in the terminal airspace. The feeder controller is responsible for providing separation and sequencing inbound aircraft toward the final approach course. The feeder controller will hand off to the arrival controller.

[8] The Arrival sector is located within the terminal airspace. The controller provides separation and sequencing of aircraft on the approach. The arrival controller will hand off to the tower controller.

training procedures, completed voice-training on a limited vocabulary, and gained familiarity and experience with the voice recognition system by working a 15 minute practice scenario on a generic airspace. Then full voice training commenced. Since it took several hours to complete voice training, the controller took several breaks while training the Verbex system on his/her voice characteristics. When completed, the controller was given a 15 minute facility-specific scenario to work during which the SME determined whether additional voice training was warranted and provided the controller with additional practice on the simulator.

Prior to beginning the experiment on Day 2, the controller once again worked a generic practice scenario. The first experimental simulation was loaded, the audio/video equipment turned on, and the controller received a standard position relief briefing from the SME. The controller used standard phraseology and followed facility procedures to provide air traffic services for aircraft during the 35-45 minute scenario. Afterwards, the controller took a break while a new scenario was loaded. This procedure was repeated until the 6 experimental scenarios were completed. A 45-minute break for lunch was provided. The following constraints were imposed on the order of scenario presentation: (1) The controller did not work 3 consecutive high traffic scenarios, (2) the controller worked traffic on each of the 3 positions before working traffic on the same position again, and (3) all controllers worked the Arrival position first.

2.6 Derived Measure of Workload

All transmissions were transcribed verbatim by a retired air traffic control specialist. An aircraft was counted as being under positive control once it established initial contact with the controller. It was no longer under positive control after the controller completed the 2-stage hand-off procedure: 1) an automated radar hand-off and 2) transfer of radio communication to the next controller in the sequence. The number of aircraft on frequency at the time the controller made a transmission was recorded next to that transmission.

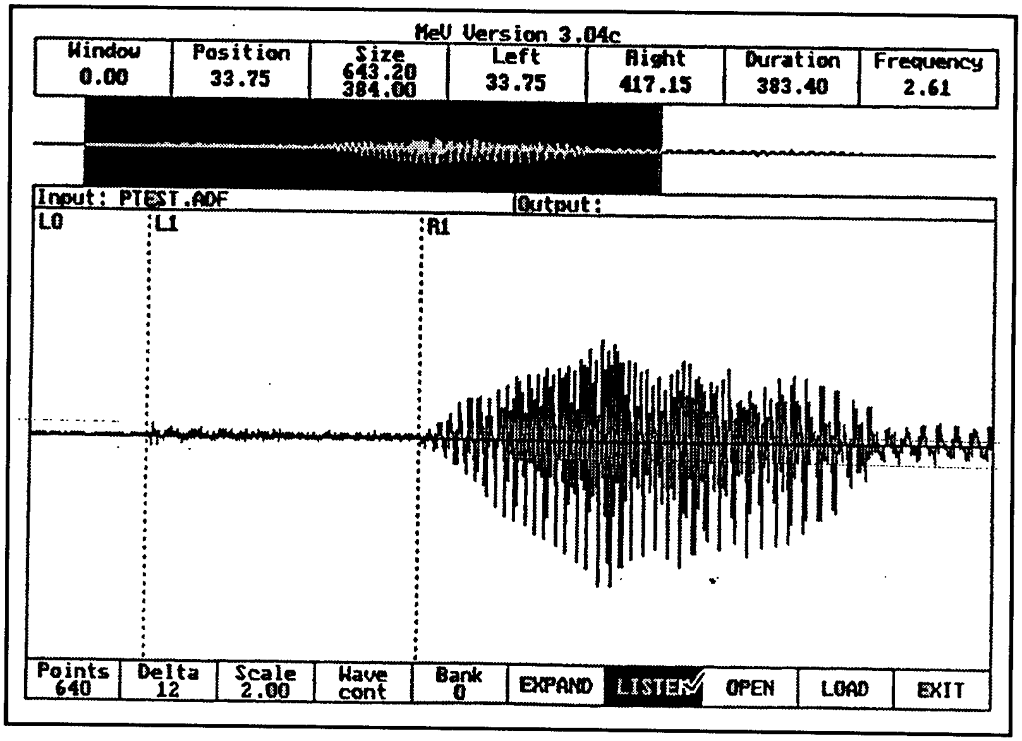
2.7 Speech Analysis Procedures

Two approaches were taken in the speech analysis, narrow and broad. First, a detailed examination of the speech of a single controller (Participant 1) was performed to look for reliable relationships within a scenario. Measures were taken from all utterances produced by this participant in the Feeder East Heavy and Feeder West Heavy scenarios. Heavy traffic scenarios were selected because they contained the most aircraft for a controller to provide radar services. They should reflect a light workload at the onset of the simulation and build to a heavy workload as the simulation progressed. Workload was determined from the total number of aircraft on frequency when the controller began speaking. East Heavy workload varied from 1-12 aircraft on frequency and West Heavy workload varied from 1-15 aircraft on frequency. Also, the increase in the number of aircraft increased the total number of transmissions available for analysis.

Second, a subset of utterances produced by the remaining participants was analyzed to assess the generalizability of Participant 1 results. Utterances were selected to achieve maximal workload contrast for each participant. For each participant, 10 utterances were identified and digitized. Those utterances corresponded to the 5 lowest workload utterances from the Light version of the scenario, simulating that participant's specialty, and the 5 highest workload utterances from the Heavy version of the scenario, simulating the participant's non-specialty sector.[9] All speech signals were sampled at 16 bits quanitization at 20,000 samples per second; the digitized signal was stored in audio files.

The analysis was performed using the interactive BLISS speech analysis system developed by John Mertus (Lieberman and Blumstein, 1988). The BLISS system permits trained operators to monitor and modify analysis parameters at virtually all stages of analysis, thereby minimizing artifacts that otherwise can be introduced by most commercially available speech analysis software. The BLISS system allows operators to view the waveform and position 4 independent sets of "cursors," e.g., left cursor L0 and right cursor R0, on the waveform. The operator can

[9] For example, for an East specialty controller, the 5 lowest workload utterances from the East Light scenario were contrasted to the 5 highest workload utterances from the West Heavy scenario.



MeV Version 3.04c

Window	Position	Size	Left	Right	Duration	Frequency
0.00	33.75	643.20 384.00	33.75	417.15	383.40	2.61

Input: PTEST.ADF

Output:

Points	Delta	Scale	Wave	Bank
640	12	2.00	cont	0

Figure 3. BLISS system display showing the waveform of the word "pig." Cursors L1 and R1 mark VOT of the initial consonant [p].

listen to the waveform delineated by any set of cursors; the sectioned waveform can be transferred to another file, "spliced" to any other file, reduplicated, scaled up or down in amplitude, inverted, etc. Figure 3 illustrates some of the features of the BLISS system. The amplitude of the speech signal at the onset of the word is displayed on the ordinate as a function of time which is plotted with respect to the abscissa.

The upper part of Figure 3 shows various aspects of the BLISS system's "header." It identifies the name the stored audio file, the cursor positions, and the waveform of the complete file, i.e., the word "pig." The lower boxes control a number of parameters of the BLISS system by means of a mouse and display the chosen parameter values. "Points" indicate the number of points that are displayed on the screen; they can be varied from 32 to 2480, allowing the operator to view and manipulate the signal with different temporal resolution. The "Delta" command instructs the system to display every Xth data point, compressing the signal. "Wave cont" is a switch that can be set to display individual data points, or as in the display of Figure 3, interpolate between data points. The "Bank" box allows 1 of 4 sets of cursors to be displayed and moved. "Expand" transfers the reverse-field (black background) display from the upper waveform display to the full screen. "Listen" allows the operator to listen to the section between any set of cursors on the total waveform displayed above. The "open" and "load" boxes are used to open new files and to transfer data to these files, e.g., the waveform between any set of displayed cursors.

For all participants, speaking rate, pause frequency, and pause duration were generated from the utterances selected as described above. **Speaking rate** (syllables/second) was computed from the number of syllables per utterance and utterance duration in ms. Because elisions and contractions (e.g., "merican" for "American") were common, only syllables actually uttered, as determined by listening to the speech sample and by visual examination of the waveform, were counted, rather than number of syllables prescribed by standard English pronunciation. Utterance duration was measured by placing cursors at the onset and offset of speech as determined by visual examination of the waveform and by listening to the speech sample. **Pause frequency** (number of pauses/ number of words) was computed from the number of pauses per utterance and the number of words per utterance. The pauses in speech are normally of too short a duration to be auditorily perceptible. Thus, pauses were identified by visual examination of the waveform. A pause was defined to be a "flat" portion of the waveform greater than 25 ms. Some articulatory gestures of speech (e.g., stop consonants) necessarily result in brief periods of silence. Using a 25 ms. lower bound excludes these articulatory factors. **Pause duration** (duration of pauses/number of words) was measured by placing cursors at the onset and offset of silence, as indicated by flattening of the waveform.

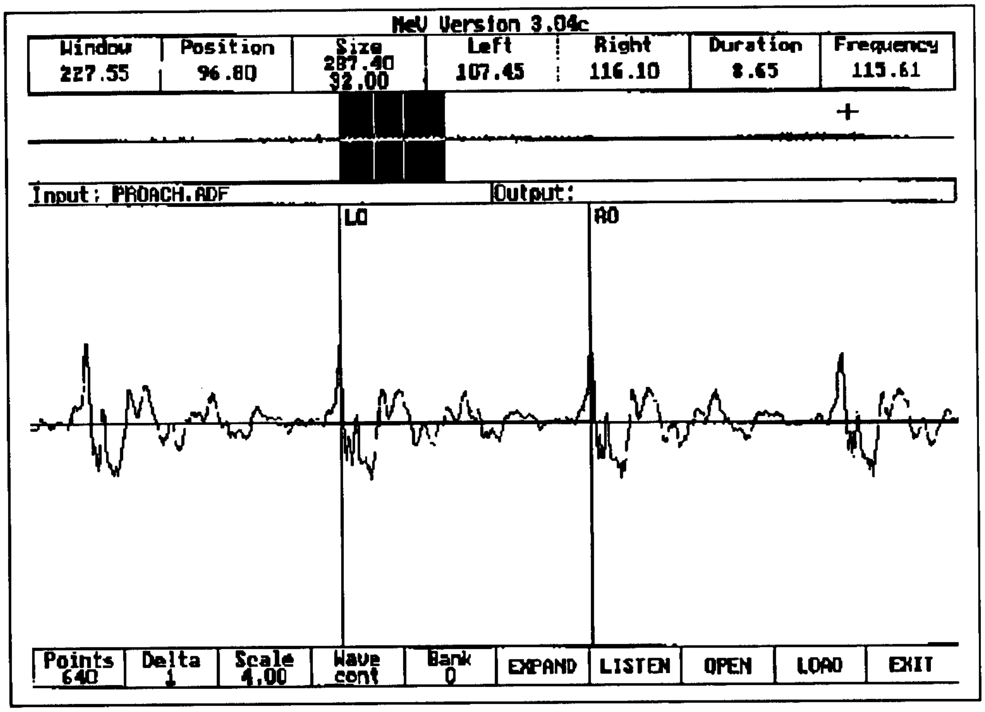
The comprehensive analysis of Participant 1's speech included 3 additional measures. For each utterance, an **F0 track**, or pitch track was computed for the entire utterance. Pitch analysis was done by use of the Short-Term Autocorrelation algorithm. Fundamental frequency is the lowest harmonic in the Fourier decomposition of a complex waveform. The Autocorrelation method extracts this harmonic from the waveform (Lieberman and Blumstein, 1988). The resulting pitch tracks were analyzed using an interleavings and offsets method, in which individual pitch tracks are interleaved and the spread, or offset, is assessed.

Additional measurements were performed on the subset of utterances that included the words "contact approach," when found in the utterance-final position (26% of utterances). This data set was chosen to minimize variation attributed to context, both lexical and phrasal. VOTs were measured for the voiceless velar stop [k] at the beginning of the word "contact" and the voiceless alveolar stop [t] at the beginning of the syllable "-tact." Cursors were placed at the onset of the burst produced at the release of the each stop consonant and at the onset of phonation, by means of both visual inspection of the waveform and by listening to marked portions of the signal. The duration of the **pitch period** (i.e., a single opening and closing of the vocal folds, as described above in Section 1.1) of the highest amplitude portion of the vowel in the syllable "-proach" is depicted in Figure 4. It was measured by placing cursors on 2 successive peaks of the waveform.

3.0 RESULTS

3.1 Narrow Focus, Participant 1

Three hundred and thirty-three utterances produced by Participant 1 make up the data set in this section (150 utterances East Heavy scenario, 183 West Heavy scenario). Reliable relationships between the utterance measures and workload for each scenario were examined. No direct statistical comparisons between East and West data were performed.



NeV Version 3.04c

Window	Position	Size	Left	Right	Duration	Frequency
227.55	96.80	287.40 92.00	107.45	116.10	8.65	115.61

Input: PROACH.ADF

Output:

Points	Delta	Scale	Wave	Bank
640	1	4.00	cont	0

Figure 4. BLISS system display showing the waveform of the syllable "proach." Cursors L0 and R0 mark a single pitch period from the high amplitude portion of the vowel.

Table 1. Narrow Focus Analysis Exploring the Relationship between Acoustic Measures and Workload for Participant 1 (East Specialty)

Dependent Measure	East Scenario	West Scenario
Speaking rate	$r = 0.038$ ($p = 0.64$)	$r = 0.095$ ($p = 0.20$)
Pause frequency	$r = -0.058$ ($p = 0.48$)	$r = -0.170$ ($p = 0.02$)
Pause duration	$r = -0.076$ ($p = 0.36$)	$r = -0.059$ ($p = 0.42$)
Velar VOT	$r = -0.581$ ($p = 0.70$)	$r = 0.112$ ($p = 0.47$)
Alveolar VOT	$r = -0.106$ ($p = 0.49$)	$r = 0.113$ ($p = 0.47$)
Pitch period	$r = -0.148$ ($p = 0.33$)	$r = -0.273$ ($p = 0.07$)

Table 2. Speech Measures Presented by Light and Heavy Workload

Dependent Measure	Light Scenario	Heavy Scenario
Speaking Rate (syl/per sec)	6.17	5.83
Standard Deviation	*1.27*	*1.25*
Pause Frequency (pauses per word)	.008	.006
Standard Deviation	*.009*	*.008*
Pause Duration (in ms)	10.47	7.39
Standard Deviation	*17.93*	*14.59*

The results of correlational analyses performed for each acoustic measure and workload are summarized in Table 1. The standard scientific convention of setting $p < .05$ was used to indicate statistically significant results. Although pause frequency and workload for the West Heavy scenario were significantly correlated, the practical significance of this result is limited. Only 3% of the variability in pause frequency was accounted for by workload. As workload increased, pause frequency decreased. Figures 5-16 (Appendix A) show average speaking rate, average pause frequency, average pause duration, velar VOT, alveolar VOT, and pitch period as a function of workload. An examination of the F0 tracks revealed variations in contour dynamics within the range noted in previous studies of single speakers (Atkinson, 1973; Lieberman et al., 1984) and were therefore concluded to be nonsignificant.

3.2 Broad Focus, Participants 2 - 12

Presented in Table 2 are summary statistics for each of the speech measures and workload. Participants 2-12 data on each of these measures are reported in Figures 17-19 and can be found in Appendix A.

3.2.1 Speaking Rate. Figure 17 shows the average speaking rate (SR) for each participant during the simulation of the Light and Heavy traffic scenarios. A visual inspection of the data reveals that overall, there appears to be a trend towards faster speech (more syllables per second) during the Light, compared with

Table 3. Broad Focus Analysis Exploring the Relationship between Mean Normalized Speech Measures and Workload

Dependent Measure	Light Scenario	Heavy Scenario
Mean Normalized Speaking Rate (SR)	*1.03*	*0.97*
Standard Deviation	(0.19)	(0.18)
Mean Normalized Pause Frequency (PF)	*1.41*	*0.59*
Standard Deviation	(*1.90*)	(*0.96*)
Mean Normalized Pause Duration (PD)	*1.46*	*0.54*
Standard Deviation	(*2.04*)	(1.20)

the Heavy traffic simulation. Sixty-four percent (7 out of 11) of the participants show this pattern, 3 show the reverse pattern, and 1 participant's SR does not change across simulations.

To determine whether the observable trend towards faster speech in the Light simulation was significant on a group basis, a normalized SR measure was generated by computing, for each participant and each SR value, the ratio of that value to the participant's mean SR. The resulting ratio values were combined into group measures and are presented in Table 3. A statistical comparison then was performed on the normalized SRs from the Light and Heavy simulations. The mean normalized SRs for the Light and Heavy simulations was not statistically significant [*t* (108) = 1.68, *p* = 0.1].

3.2.2 Pause Frequency. Figure 18 shows average pause frequency (PF) for each participant in the Light and the Heavy simulation. A visual inspection of the data reveals a trend towards more frequent pauses in the Light rather than in the Heavy simulation. Seven of the 11 controllers show this pattern, and 3 show the reverse pattern. Only Participant 11 shows constancy in average pause frequency in both traffic conditions.

To determine whether the observable trend towards more frequent pausing during the Light simulation was significant on a group basis, normalized PF measures were generated. This was accomplished by computing, for each participant and each PF value, the ratio of that value to the participant's mean PF. A statistical comparison then was performed on the normalized PFs from the Light and Heavy simulations. The mean normalized PFs for the Light and Heavy simulations were significantly different [*t* (108) = 2.86, *p* = .05]; controllers produced more pauses during the Light simulation.

3.2.3 Pause Duration. Figure 19 shows average pause duration (PD) for each participant in the Light and the Heavy simulations. Again, there appears to be a trend towards longer pause durations in the Light than in the Heavy simulation. Nine of the 11 participants show this pattern and 1 shows the reverse pattern. Participant 7 shows non-discernable variation in average pause duration.

To determine whether the observable trend towards longer pausing in the Light simulation was significant on a group basis, normalized PD measures were generated. This was accomplished by computing, for each participant and each PD value, the ratio of that value to the participant's mean PD. A statistical comparison then was performed on the normalized PDs from the Light and Heavy simulations (1.46 and 0.54, respectively). The difference between the mean normalized PDs was significant [*t* (108) = 2.85, *p* = .05], controllers paused longer during the Light simulation.

4.0 DISCUSSION

The data for this set of analyses contrasted acoustic measures of communication generated by controllers while they provided radar services to pilots on the sector of their specialty (light traffic) and on a sector other than their specialty (heavy traffic). All participants were full-performance journeymen controllers who were highly skilled and knowledgeable about their airspace and procedures. Workload was measured as the number of active aircraft on frequency at the moment a controller initiated an utterance. To maximize the likelihood of significant results between workload and the selected acoustic measures, the 5 utterances transmitted under the lightest and heaviest workload simulations were examined for 11 of the 12 controllers. A very detailed and complex set of acoustic analyses was performed only on the data from 1 controller.

The results presented here suggest 2 points of interest. First, as a group, the ATCSs who participated in this simulation study displayed a tendency both to pause more frequently and pause longer during a light rather than heavier workload situation. From these results, it is possible to infer that the type of "hesitation" produced by this group of controllers is not associated with factors such as task difficulty, as described above in Section 3. Instead, these data may reflect the possibility that, when workload is light, controllers may attend to the task in hand using a "cognitive" rather than an "automatic" response mode. Under a light traffic load, controllers had more flexibility and latitude in determining runway assignments and sequencing aircraft for the approach. Light traffic coupled with the expertise of working on their own sector specialty allowed for more thinking time, especially when the constraints imposed by rapidly converging aircraft into a small airspace were removed.

At busy Level 5 TRACON facilities, standard terminal approaches are used, and pilots know that at particular locations they must have their aircraft at a particular altitude, heading, and airspeed. As part of their training, controllers learn when to descend, slow, turn, and clear an aircraft for an approach; they also learn when and how to transfer radar and radio communication to the next controller in the sequence. Since they deliver this information over and over again, hesitations diminish. Under heavy traffic, a more highly automatic, routinized approach to traffic management became operational, and communication with pilots became "canned," and repetitive. Under periods of heavy workload, more routinized cognitive processes might occur, as demonstrated by fewer pauses of shorter duration. This possibility is likely in light of previous studies of pause duration and frequency (Eisler 1968) and given the lack of a statistically significant change in speaking rate.

It is especially interesting that both measures of hesitation increased. The measures used are, in principle, independent of one another. That is, because pause duration is averaged across all pauses in an utterance, there is no *a priori* reason to suppose that an increase in the number of pauses would be associated with an increase in the length of those pauses. And, in fact, in some cases ATCSs who showed the dominant trend on one of these measures showed the opposite of the dominant trend on the other measure. Only 55% participants showed both trends, as displayed in Figures 18 - 19.

The Light simulation utterances were on average slightly longer in duration (3218 ms vs. 3027 ms) and slightly greater in number of syllables spoken (18.6 vs. 16.6). Although these differences were not statistically significant, it may be the case that "more speech" provides more opportunity both for more frequent and longer pauses. However, the occurrence of both longer Type 2 normalized pause durations and normalized pause frequencies is consistent with the controllers responding in a more cognitive mode under the Light condition.

The second point of interest is that while there were strong group effects for the 2 measures of hesitation, these effects were rarely significant on an individual basis. Further, regardless of the size of the effects, none of the 3 measures showed trends in the same direction for all participants. This was especially true for speaking rate, for which 2 participants showed significant differences in opposite directions.

The importance of this fact, the variability in speaking among participants, is highlighted by the results from the more in-depth analysis of the speech of Participant 1. Despite the analysis of more than 300 utterances, only 1 significant, albeit weak correlation was found between workload and the many acoustic measures. We conclude this report by suggesting that the data from Participant 1 may not be representative of the pool of data provided by the other 11 participants.

The results of analyses performed for each acoustic measure and workload lead us to conclude that hesitations found in speech may be a potential indicator of workload, as measured by pause duration and pause frequency, in particular. Despite its highly speaker-dependent nature, hesitation may prove to be a useful indicator of a controller's responding in a cognitive, rather than in an automatic mode. The exhaustive data collected by Eisler and her colleagues show that individuals who are devoting fewer cognitive resources to a discussion manifest shorter Type 2 pause durations than people thinking about what they are communicating. The speech of the controllers in this study, therefore, may reflect a shift between a more cognitive "thinking" response mode in Light traffic situations, where they know that they have more time to respond, and a more automatic mode that allows them to respond to the increased pace induced by higher traffic loads. In other words, we may be monitoring the degree to which the controllers respond by means of reasoned, cognitive rather than automatic, routinized responses.

Our research suggests that future investigation of the speech of controllers should include measures of hesitation as a measure of interest. Also, when considering potential acoustic-phonetic candidates for inclusion in speech-derived measures of workload, individual differences must be considered because of the demonstrated variably in speaking among the participants in this study. Thus, the results presented in this report indicate that the examined speech measures do not generalize across people and should not be used to make inferences about groups. Speech production is highly individualized and varies with the speaker and situation.

5.0 REFERENCES

Absil, E., Grammatica, B., Harmegnies, B., Legros, C., Poch, D., and Ruiz, R. (1995). Time related variabilities in stressed speech under laboratory and real conditions. *Proceedings of the ESCA - NATO Tutorial and Research Workshop on Speech under Stress,* (pp. 53-56). Portugal: Colibri, Sociedade de Artes Graficas.

Atkinson, J. (1973). Aspects of intonation in speech: Implications from an experimental study in voice fundamental frequency. Ph.D. Dissertation, University of Connecticut.

Baum, S.R., Blumstein, S.E., Naeser, M.A., and Palumbo, C.L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language, 37,* 327-338.

Benson, P. (1995). Analysis of the acoustic correlates of stress from an operational aviation emergency. *Proceedings of the ESCA - NATO Tutorial and Research Workshop on under Stress,* (pp. 61-64). Portugal: Colibri, Sociedade de Artes Graficas.

Blumstein, S.E., Cooper, W., Goodglass, H. Statlender, H. and Gottleib, J. (1980). Production deficits in aphasia: a voice-onset time analysis. *Brain and Language, 9,* 153-170.

Borden, G.J., and Harris, K.S. (1984). Speech Science Primer: Physiology, Acoustics, and Perception of Speech, Second Edition. Baltimore: Williams and Wilkins.

Coster, W.J. (1986). *Aspects of voice and conversation in behaviorally inhibited and uninhibited children.* Unpublished Ph.D. dissertation. (Harvard University Archives HU90. 12239 Harvard Depository).

Cummings, K.E., and Clements, M.A. (1990). Analysis of glottal waveforms across stress styles. *Proceedings IEEE ICASSP (IEEE International Conference on Acoustics, Speech, and Signal Processing),* (pp. 369-372). Piscatawya, NJ: IEEE Service Center.

Eisler, F.G. (1968). *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Frick, R.W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97,* 419-429.

Kagan, J., Reznick, J.S., and Snidman, N. (1988) Biological bases of childhood shyness. *Science, 240,* 167-171.

Kessinger, R., and Blumstein, S.E. (in submission). *Rate of speech effects on voice onset time in Thai, French, and English.*

Ladefoged, P. (1962). *Elements of acoustic phonetics.* Chicago: The University of Chicago Press.

Lieberman, P. (1967). *Intonation, perception and language.* Cambridge: MIT Press.

Lieberman, P. (1963). Some measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America, 35,* 344-353.

Lieberman, P., and Blumstein, S.E. (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge: Cambridge University Press.

Lieberman, P., and Michaels, S.B. (1962). Some aspects of fundamental frequency, envelope amplitude and the emotional content of speech. *Journal of the Acoustical Society of America, 34,* 922-927.

Lieberman, P., Protopapas, A. and Kanki, B.G. (1995). Speech production and comprehension deficits on Mt. Everest. *Aviation, Space, and Environmental Medicine, 66,* 857-869.

Lieberman, P., Kako, E.T., Friedman, J., Tajchman, G., Felldman, L.S., and Jiminez, E.B. (1992). Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language, 43,* 169-189.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., and Miller, M. (1984). Measures of the sentence intonation of read and spontaneous speech in American English. *Journal of the Acoustical society of America, 77,* 649-657.

Lisker, L. and Abramson, A.S. (1964). A cross language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-342.

Muller, J. (1848). *The physiology of the senses, voice and muscular motion with the mental faculties.* (W. Baly, Trans.). London: Walton and Maberly.

Prinzo, O.V. (1996). *An analysis of approach control/ pilot voice communications.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-96/26, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA274457.

Prinzo, O.V. & Britton, T.W. (1993). *ATC/pilot voice communications: A survey of the literature.* Federal Aviation Administration, Office of Aviation Medicine Technical Report DOT/FAA/AM-93/20, Washington, DC. Available from: National Technical Information Service, Springfield, VA 22161; ordering no. ADA317528.

Sataloff, R.T. (1992). The human Voice. *Scientific American, 267,* 108-115.

Waters, J., Nunn, S. Gillcrist, B. and VonColln, E. (1995). The effect of stress on the glottal pulse. *Proceedings, ESCA - NATO Tutorial and Research Workshop on Speech under Stress,* pp. 9-11. Portugal: Colibri, Sociedade de Artes Graficas.

APPENDIX A

FIGURES 5-19

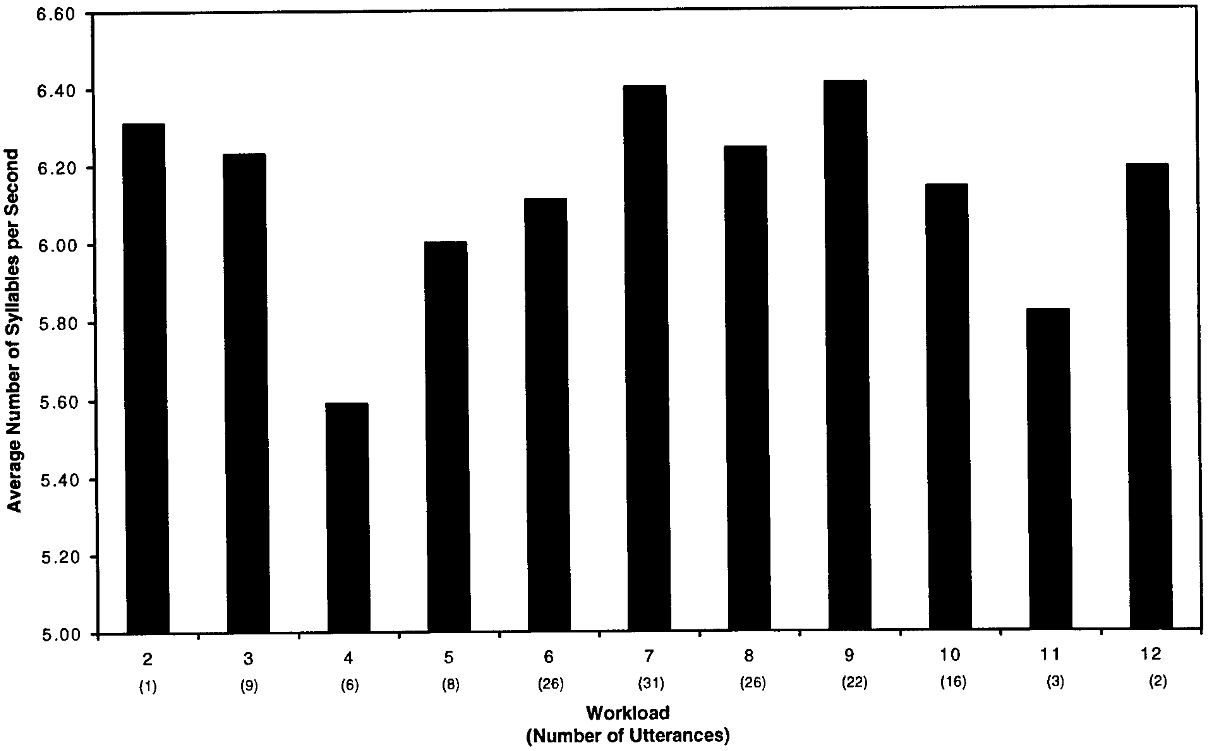


Figure 5. Average Speaking Rate (Syllables/Second) as a Function of Workload for Participant 1 in the East Heavy Scenario.

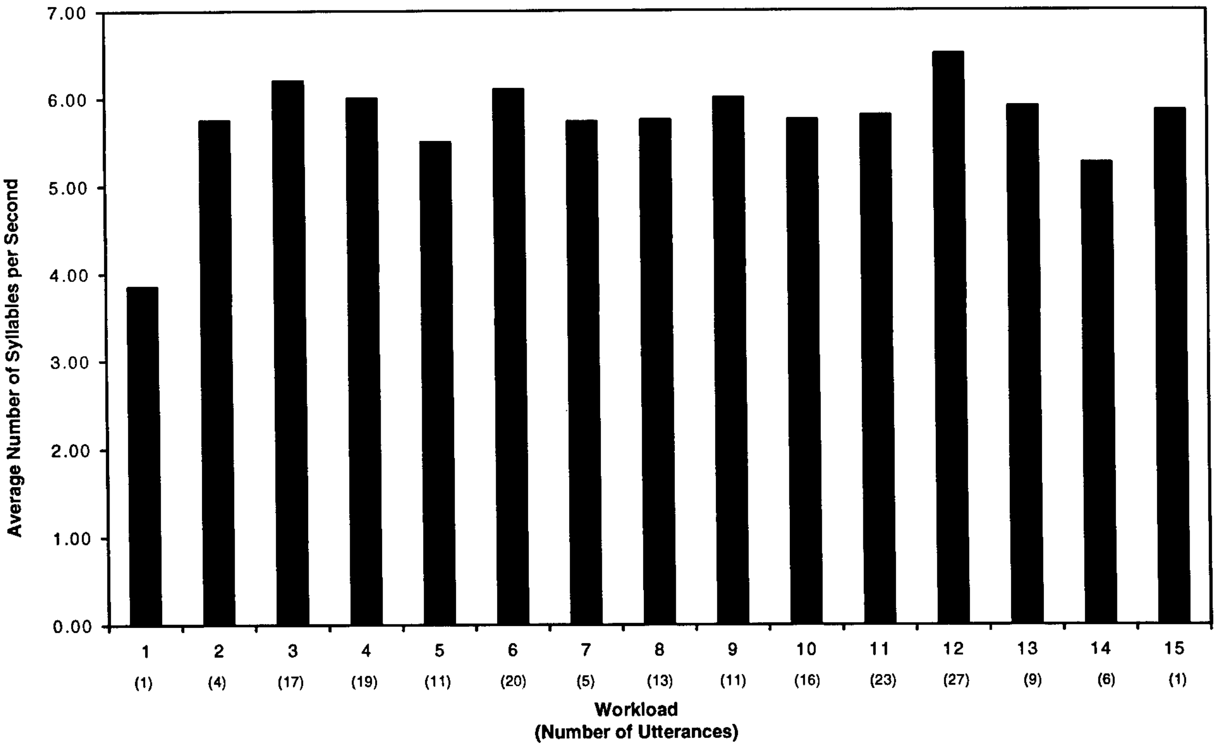


Figure 6. Average Speaking Rate (Syllables/Second) as a Function of Workload for Participant 1 in the West Heavy Scenario.

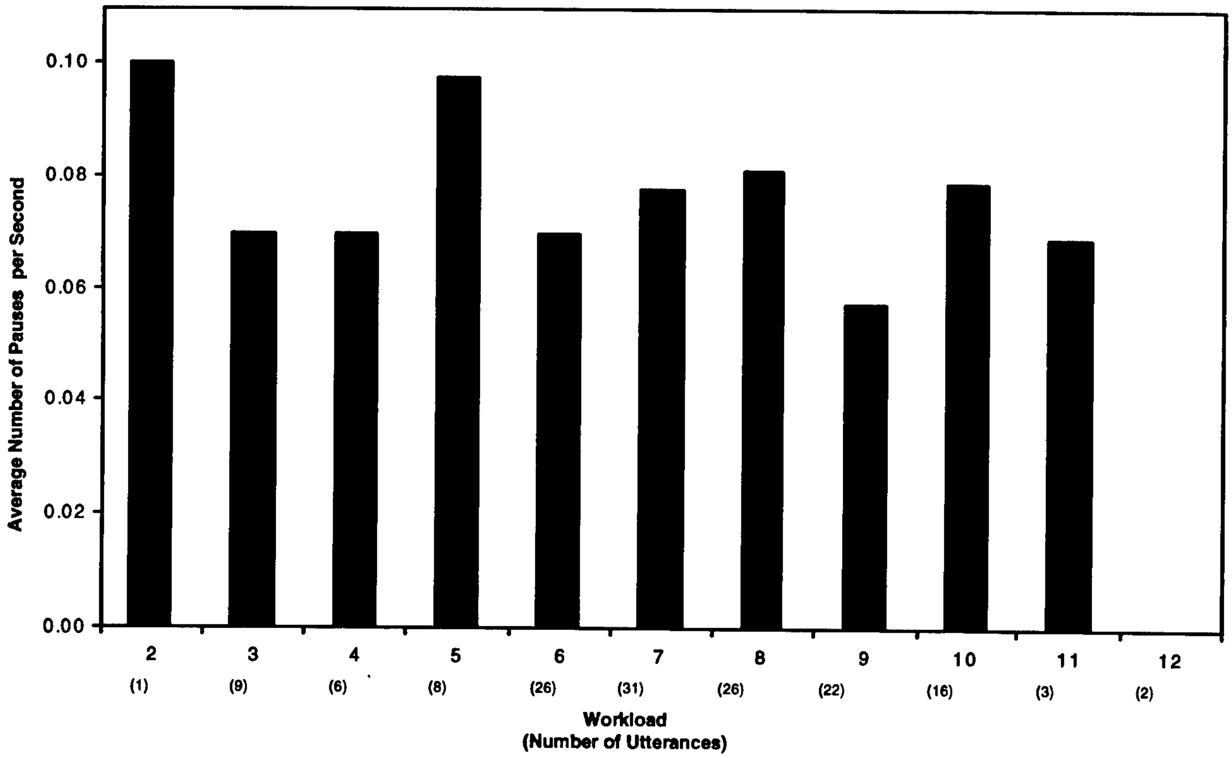


Figure 7. Average Pause Frequency (Number of Pauses/Number of Words) as a Function of Workload for Participant 1 in the East Heavy Scenario.

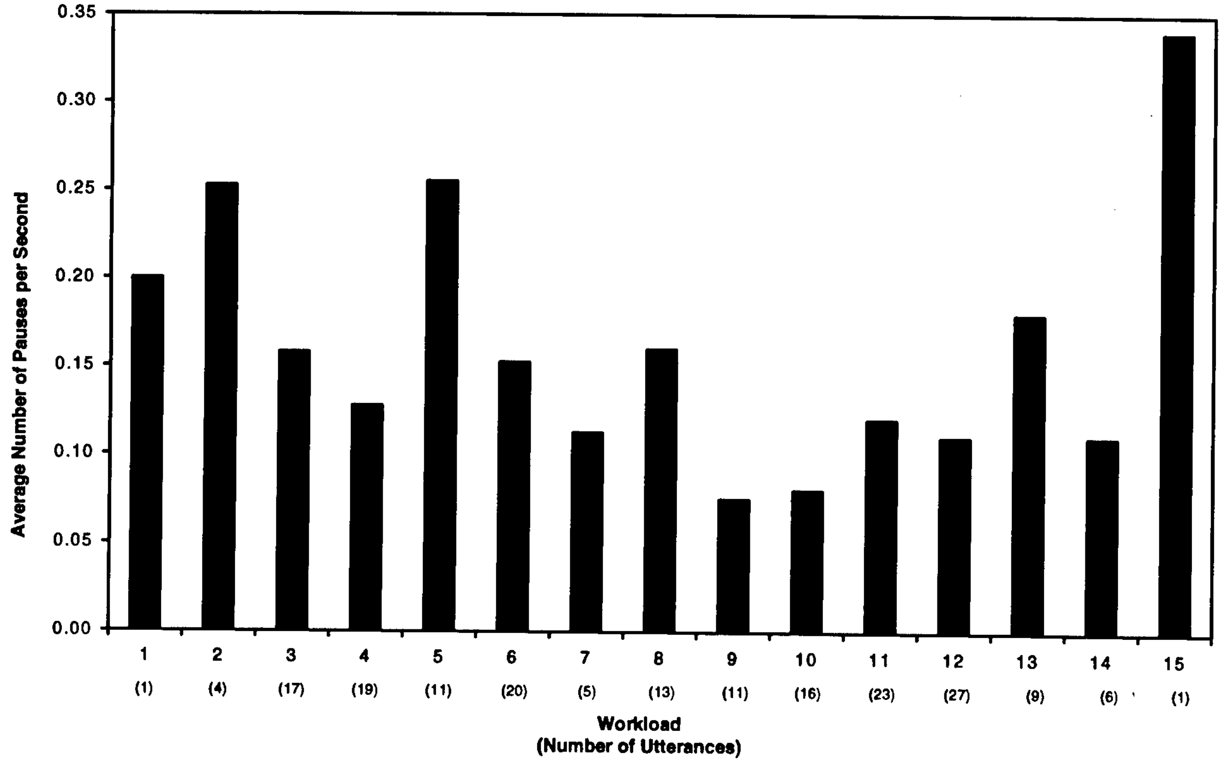


Figure 8. Average Pause Frequency (Number of Pauses/Number of Words) as a Function of Workload for Participant 1 in the West Heavy Scenario.

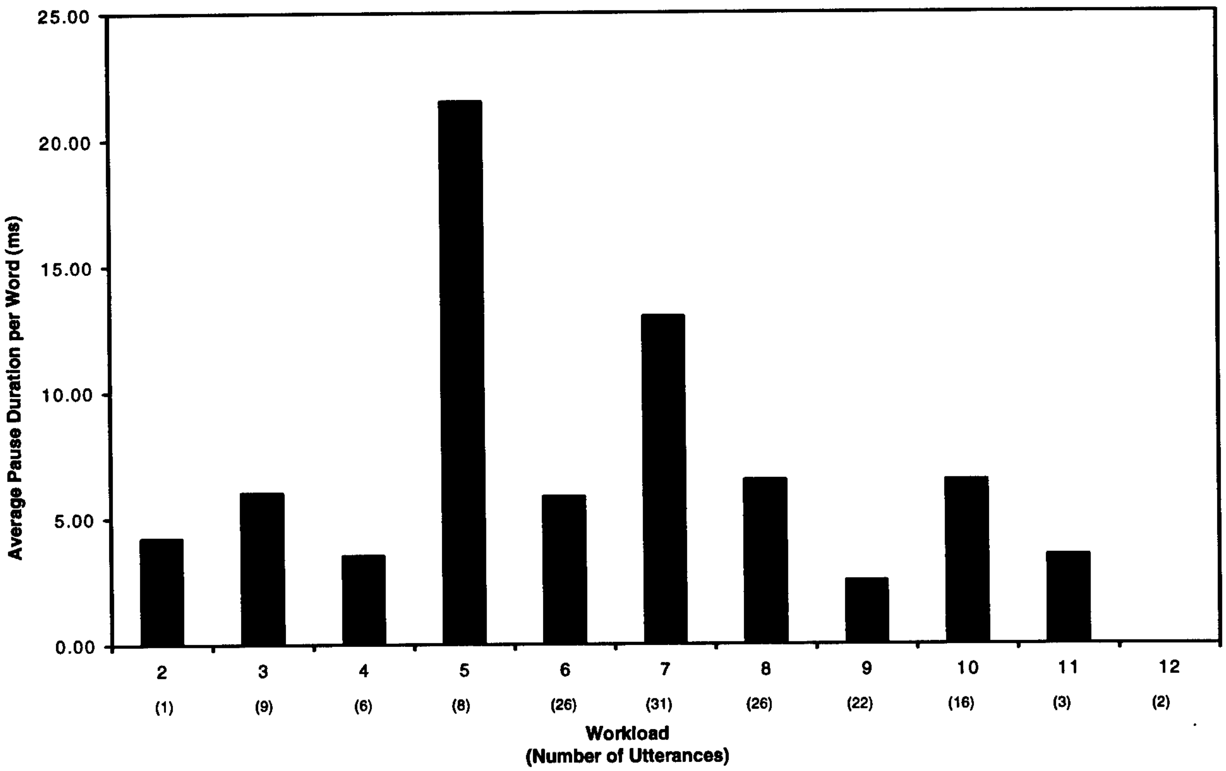


Figure 9. Average Pause Duration (Duration of Pauses/Number of Words) as a Function of Workload for Participant 1 in the East Heavy Scenario.

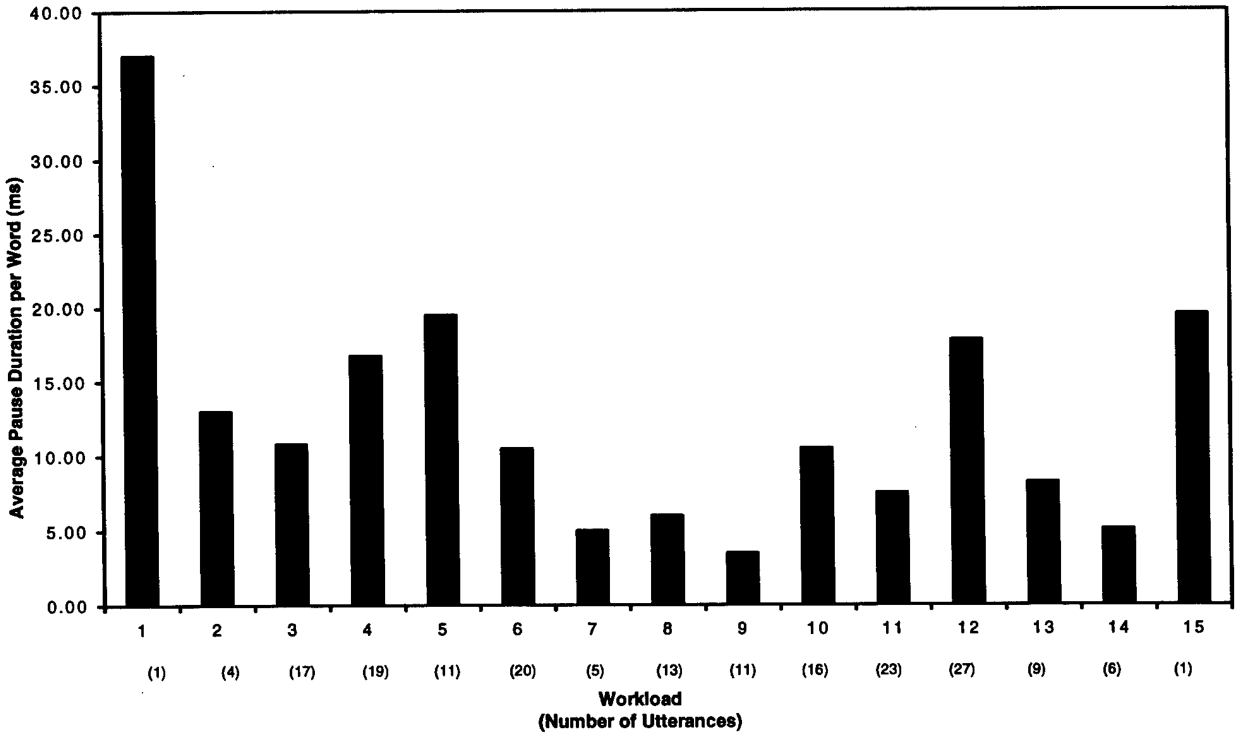


Figure 10. Average Pause Duration (Duration of Pauses/Number of Words) as a Function of Workload for Participant 1 in the West Heavy Scenario.

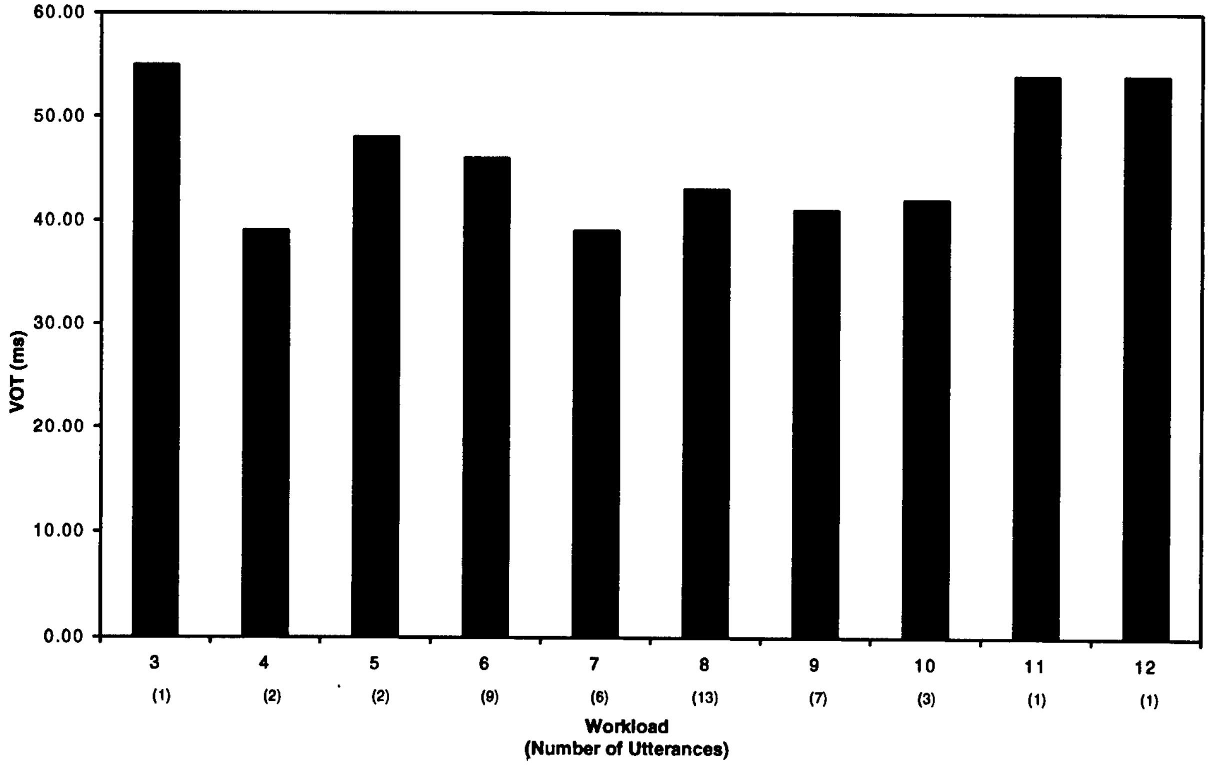


Figure 11. Average VOT of [k] as a Function of Workload for Participant 1 in the East Heavy Scenario.

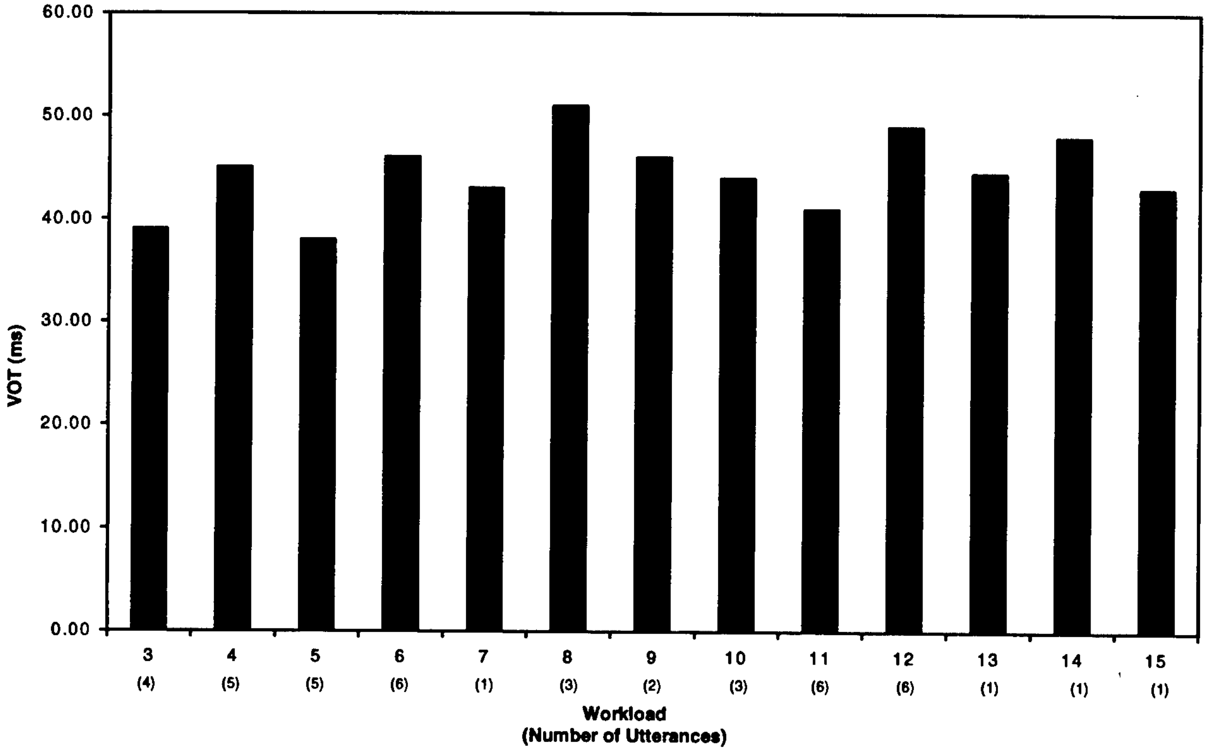


Figure 12. Average VOT of [k] as a Function of Workload for Participant 1 in the West Heavy Scenario.

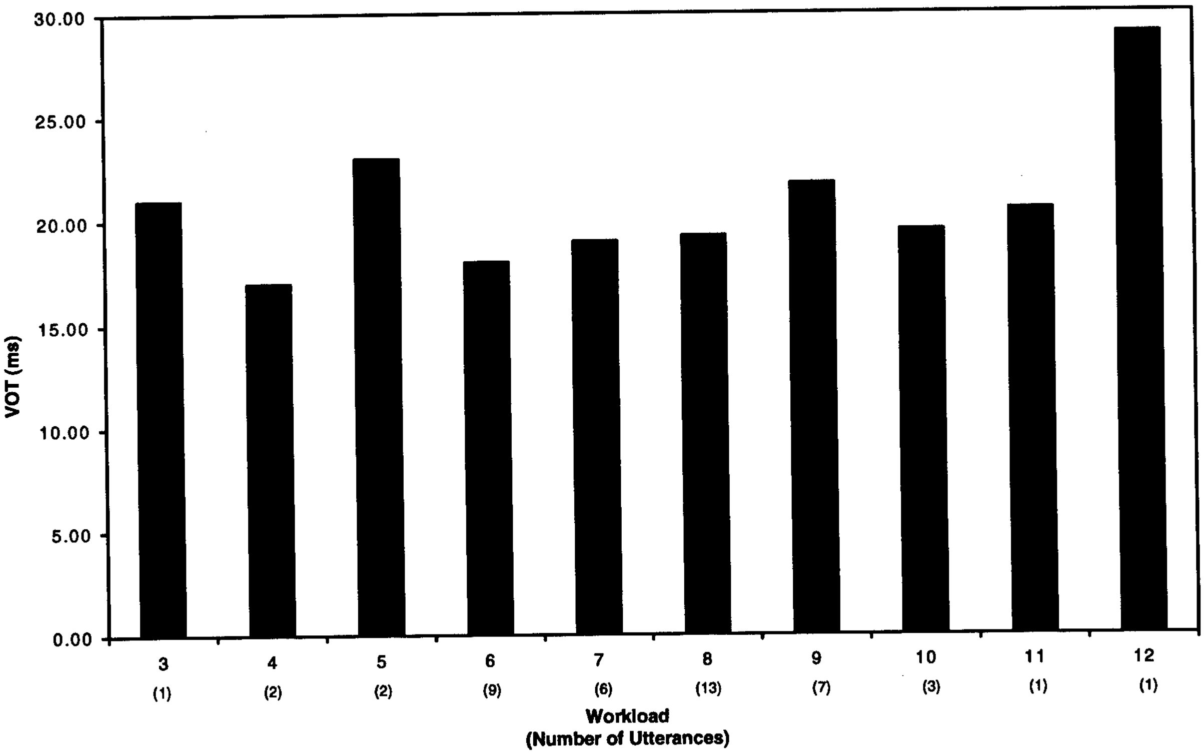


Figure 13. Average VOT of [t] as a Function of Workload for Participant 1 in the East Heavy Scenario.

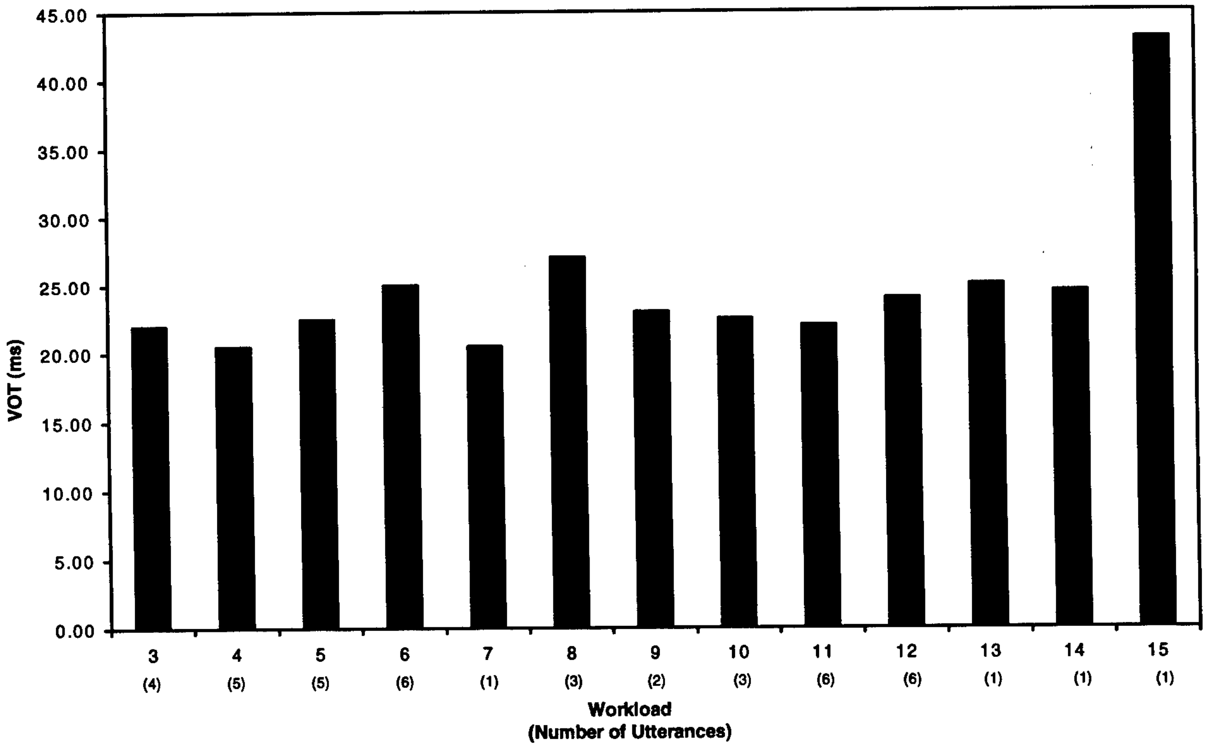


Figure 14. Average VOT of [t] as a Function of Workload for Participant 1 in the West Heavy Scenario.

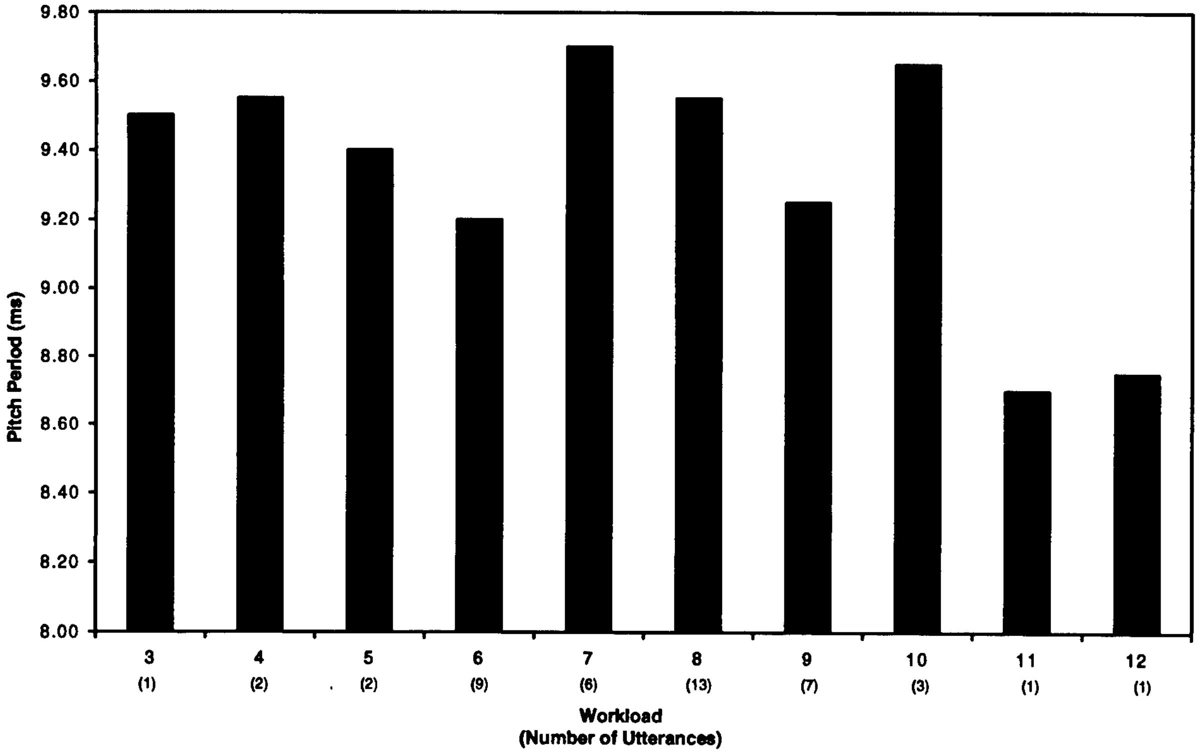


Figure 15. Average Pitch Period from Phrase Final Syllable "proach" as a Function of Workload for Participant 1 in the East Heavy Scenario.

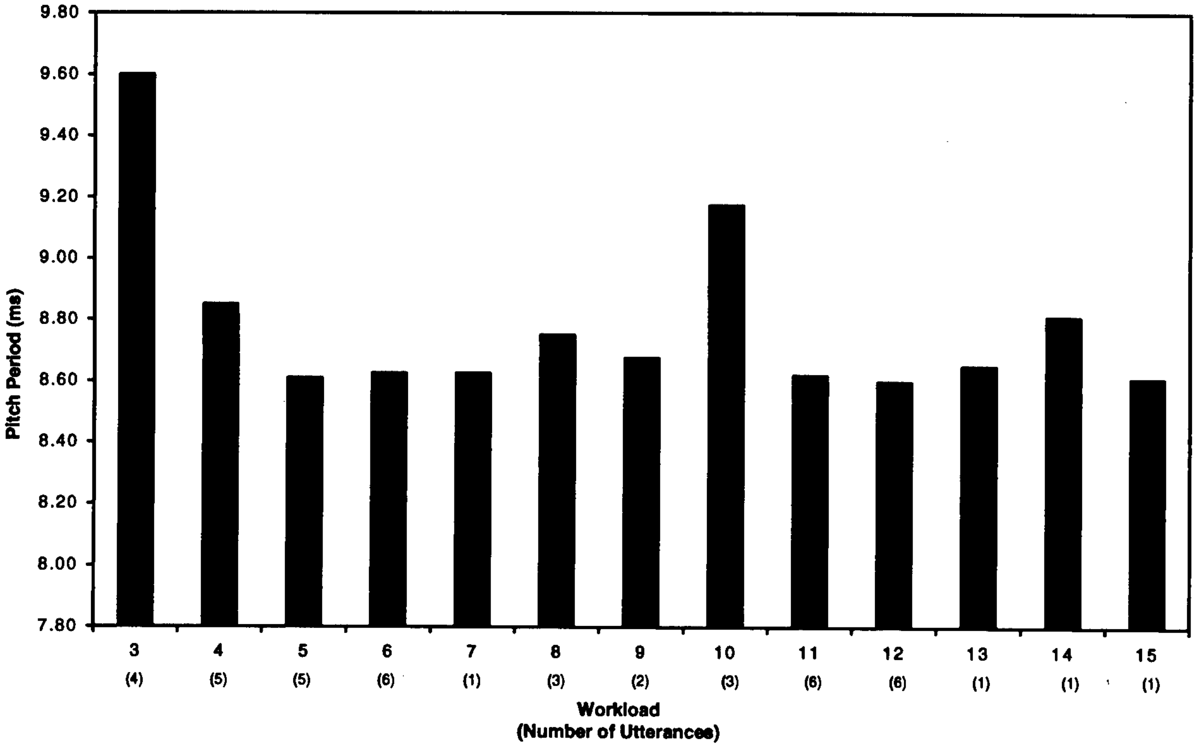


Figure 16. Average Pitch Period from Phrase Final Syllable "proach" as a Function of Workload for Participant 1 in the West Heavy Scenario.

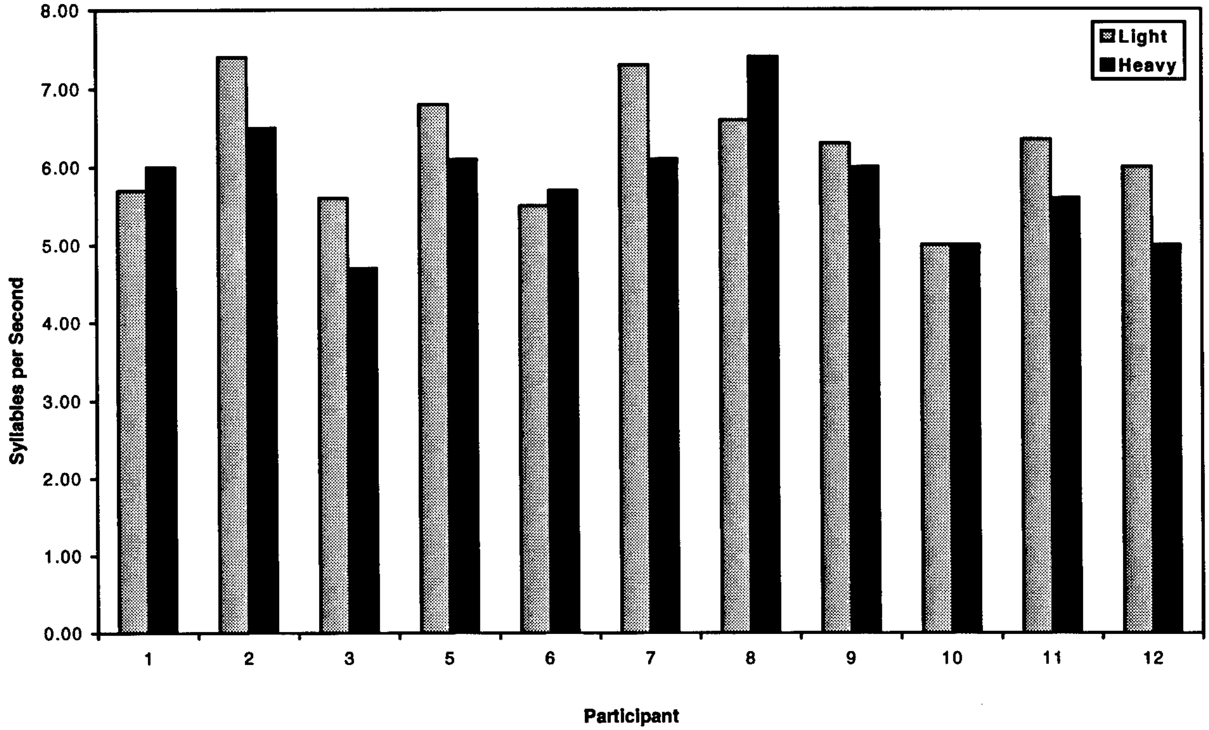


Figure 17. Average Speaking Rate (SR) for Each Participant in the Light and the Heavy Conditions.

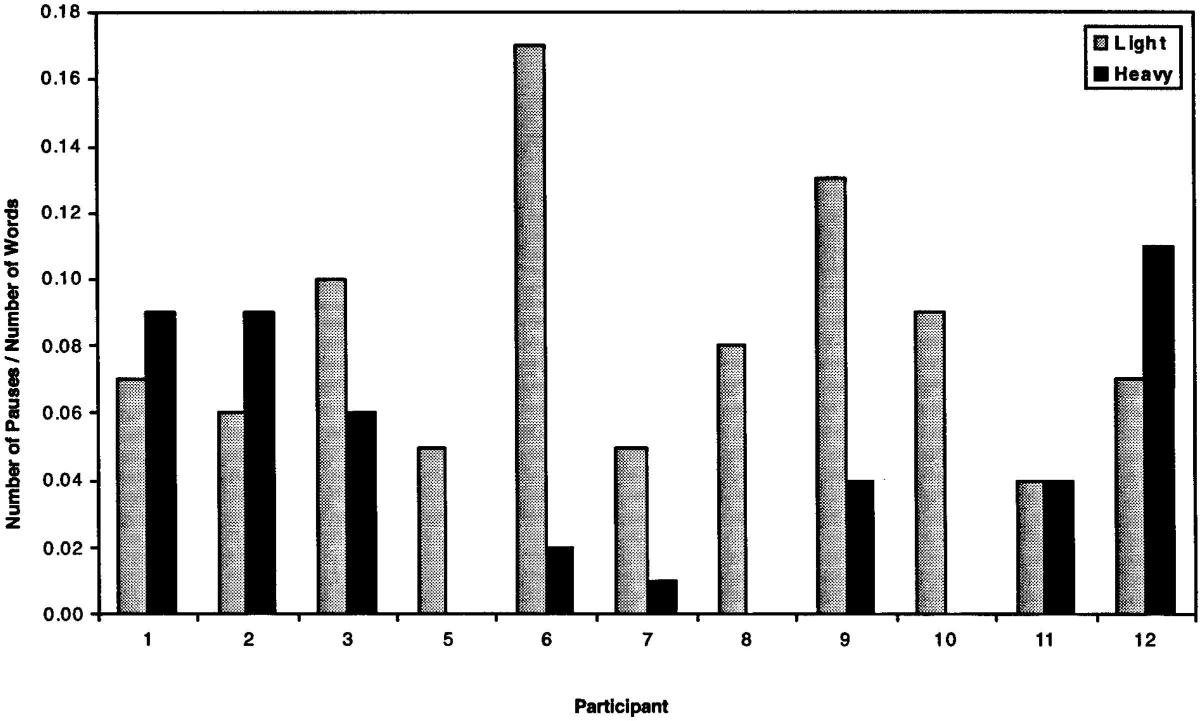


Figure 18. Average Pause Frequency for Each Participant in the Light and the Heavy Conditions.

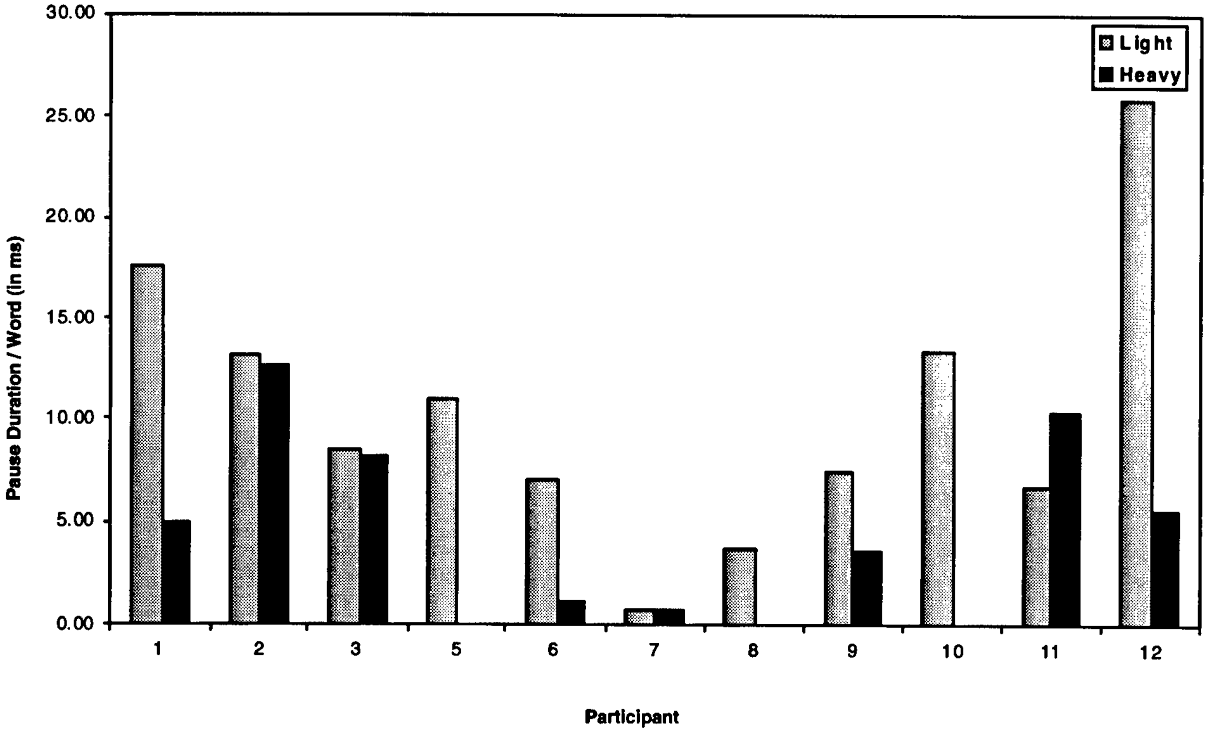


Figure 19. Average Pause Duration for Each Participant in the Light and the Heavy Conditions.

☆U.S. GOVERNMENT PRINTING OFFICE: 1998 -660-088/80004